Journal of Protein Chemistry, Vol. 11, No. 3, 1992

Sequence and Structural Relationships in the Cytokine Family

Parthasarathy Manavalan,[1,2] Deborah L. Swope,[1] and Raymond M. Withy[1]

Received January 2, 1992

The sequences of nine different cytokines, growth hormone, and prolactin have been aligned and their secondary structure predicted. The alignment reveals that each exon has a characteristic sequence pattern shared by all cytokines. The most striking sequence similarity is observed in exon 4, where the residue pair Phe-Leu is conserved in many cytokines. In addition, there are discreet homologous regions between two specific growth factors, including a high degree of homology between granulocyte-macrophage colony-stimulating factor (GM-CSF) and interleukin 3 (IL-3). The secondary structure analysis predicts that exon 3 of all cytokines has an antiparallel helix-turn-helix motif, which is likely to form the central helical segments of a four α-helical bundle-type structure. Based on the secondary structure and the disulfide-bonding pattern, the topological connectivity for a number of cytokines has been predicted.

KEY WORDS: Cytokines; sequence homology; secondary structure; folding topology.

1. INTRODUCTION

Cytokines and growth factors are regulatory polypeptides produced by cells during embryonic development, active developmental processes, such as hematopoiesis and wound healing, and during an immune response (Paul, 1989; Arai *et al.*, 1990). Included in this group of proteins are the interleukins,[3] the colony-stimulating factors, erythropoietin, growth hormone, and prolactin. These proteins control a wide range of functions in cells in the lymphoid, hemopoietic, and reticuloendothelial systems by binding to specific cell-surface receptors on target cells. Interestingly, most of the cytokines are pleiotropic and have multiple biological activities (Metcalf, 1989; Mizel, 1989; Nicola, 1989), and this functional redundancy might suggest that these proteins also have overlapping sequence and structural features.

Recently, it has been shown that the receptors for the cytokines show some interesting homologies. The amino acid sequence analysis of the extracellular domains of cytokine receptors reveals two distinctive sequence motifs suggesting that they all belong to a cytokine receptor super family (Bazan, 1989; D'Andrea *et al.*, 1989; Goodwin *et al.*, 1990; Nicola and Metcalf, 1991). Sequence homologies among cytokine ligands, unlike their receptors, are not apparent except that IL-6 shows limited homology with G-CSF and cMGF (Leutz *et al.*, 1989). However, the genomic organization of these ligands reveals that many of them have four exons encoding the mature form of the protein, irrespective of the variation in their sequence lengths, suggesting the idea that these cytokines might be related.

In an attempt to characterize the homologies and other common structural features among the cytokine ligands, we have made a detailed sequence and structural analysis of these ligands. Our sequence homology study reveals a number of discrete highly conserved regions that are unique among these growth factors. The alignment of secondary structure also indicates two contiguous α-helical segments separated by a small loop as a common structural pattern shared by all growth factors. Using this structural pattern and known disulfide bonds that connect

[1] Genzyme Corporation, Framingham, Massachusetts 01901.
[2] To whom all correspondence should be addressed.
[3] Abbreviations used: interleukin 1, IL-1; interleukin 2, IL-2; interleukin 3, IL-3; interleukin 4, IL-4; interleukin 5, IL-5; interleukin 6, IL-6; interleukin 7, IL-7; interleukin 9, IL-9; interleukin 11, IL-11; erythropoietin, EPO; granulocyte colony-stimulating factor, GCSF; granulocyte-macrophage colony-stimulating factor, GM-CSF; cMGF, chicken myelomonocytic growth factor; growth hormone, GH; prolactin, PRL; circular dichroism, CD.

0277-8033/92/0600-0321$06.50/0

different parts of the molecule, the folding topologies for a number of growth factors are predicted.

2. METHODS

2.1. Sequence Analysis

The proteins included for our structural analysis are interleukin 2 (IL-2) (Taniguchi *et al.*, 1983; Kashima *et al.*, 1985), IL-3 (Fung *et al.*, 1984; Yang *et al.*, 1986), IL-4 (Lee *et al.*, 1986; Yokota *et al.*, 1986), IL-5 (Azuma *et al.*, 1986; Kinashi *et al.*, 1986), IL-6 (Hirano *et al.*, 1986; Van Snick *et al.*, 1988), IL-7 (Goodwin *et al.*, 1989; Lupton *et al.*, 1990), IL-9 (Yang *et al.*, 1989), erythropoietin (Jacobs *et al.*, 1985; Lin *et al.*, 1985; McDonald *et al.*, 1986) (EPO), granulocyte-colony stimulating factor (Nagata *et al.*, 1986; Tsuchiya *et al.*, 1986) (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF) (Gough *et al.*, 1984; Wong *et al.*, 1985), growth hormone (GH) (DeNoto *et al.*, 1981; Page *et al.*, 1981), chicken myelomonocytic growth factor (c-MGF) (Leutz *et al.*, 1989), and prolactin (PRL) (Bartá *et al.*, 1981; Cooke *et al.*, 1981). Both human and murine sequences were examined for these proteins. Sequences were obtained from the literature and analyzed for sequence homologies using pattern-matching and structure prediction methods. Peptide sequences were initially aligned using the programs "ALIGN," based on the Needleman and Wunsch homology algorithm (Intelligenetics), and "BESTFIT" (Devereux *et al.*, 1984) (University of Wisconsin GCG), which employs the local homology algorithm of Smith and Waterman. The sequences were aligned using gap penalties of two, three, and four to maximize the local homologous regions and identify the conserved residues. Using these local homologous regions as the starting point, a multiple-sequence alignment using all peptides was constructed manually for exons 1, 2, 3, and 4 within each protein. Homology was determined by maximizing identical and nearly identical amino acids.

2.2. Secondary Structure Prediction

Secondary structure was predicted by a joint algorithm using Garnier–Osguthorpe–Robson (Garnier *et al.*, 1978), Chou and Fasman (1978), Lim (1974), and Cohen *et al* (1986) methods. Since each protein has mouse and human versions, a consensus secondary structure was determined using the following strategies.

First, the conformational parameters (α-helix, β-sheet, β-turns, and coil) for each residue was calculated for both human and murine analogs by GOR method. The values were averaged for all the aligned residues at each position. The conformational state that showed the highest value at a particular position was predicted to be the likely structure for the residue at that position. The same procedure was used to predict the secondary structure by the Chou and Fasman (C&F) method. β-turns and other irregular regions were also determined by the pattern-matching method developed by Cohen *et al.* (1986). From these predicted structures, the final assignment of secondary structure was made using the following rules.

A region was considered in β-turn or irregular structure, if two out of three methods predicted β-turns or irregular conformation for that region.

A region was considered in α-helix (or β-sheet), if both GOR and C&F methods predicted α-helix (or β-sheet) for that region [at least six (four) consecutive residues should be in α-helical (β-sheet) conformation]. Whenever there was an ambiguity in assigning a particular structure, Lim's residue distribution method (Lim, 1974) was used to determine the most probable structure for that region.

3. RESULTS

3.1. Sequence Homologies

Figure 1 shows the alignment of all growth factors included in our analysis. The most striking feature of this alignment is the location of intron/exon boundaries for each growth factor. The exon distribution for the cytokines, growth hormone, and prolactin are given in Table I. It is obvious that exon size varies depending on the size of the protein molecule. However, according to our alignment shown in Fig. 1, the exon/intron position for each growth factor falls within a narrow range irrespective of its size. The strongest alignment of exon boundaries occurs after the first exon and before the fourth exon (except HIL-6) in each protein. The location of the intron between the second and third exons is more variable. The intron/exon positions for IL-9 are not known. Based on the sequence alignment, we predict that each boundary for IL-9 should fall within the range specified by other cytokines.

3.2. Homologies in Each Exon

Figure 1 also reveals that each exon has some characteristic sequence pattern that is highly conserved in all the growth factors. The major region of

Fig. 1. Comparison of the primary sequences of cytokines and other related growth factors. Identical amino acid residues present in eight or more of the growth factor sequences are highlighted with dark background (reverse video). Highly acceptable substitutions present in more than 12 growth factors are shaded. Residues with similar properties are boxed. Highly acceptable substitutions are group 1: (L, I, V); group 2: (D, E); group 3: (R, H, K); group 4: (F, Y, W); group 5: (S, T); and group 6: (Q, N). Residues with similar properties are L, I, V, M, A, F, Y, W, C: R, K, H, Q, N. Arrows indicate intron/exon boundaries for each growth factor. Regions 1, 2, 3A, 3B, and 4 show the consensus sequence patterns. The exon lengths that are common for all growth factors are shown above the sequence alignment (broken line). The gap between two exons defines the range where intron/exon boundaries fall. Asterisks in exon 1 of HGCSF and in exon 1 of MIL2 indicate the omission of few residues (3 residues in HGCSF and 16 residues in MIL2). Few C-terminal residues are also not shown for MIL2, IL3, IL7, and IL9. A, Ala; D, Asp; C, Cys; E, Glu; F, Phe; G, Gly; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

homology between the growth factors occurs in exon 4, which contains a highly conserved aromatic residue followed by Leu (except one). This Phe-Leu pair is within 20 residues (except murine IL-3) from the C-terminal end for all the proteins included in our analysis (region 4 in Fig. 1). The aromatic residue is a phe for 19 out of 25 cytokines and the remaining six cytokines have Tyr at this position. This pair of residues is also conserved in all the species of growth hormone, prolactin, and other prolactin-related molecules, proliferin and somatolactin (Linzer and Nathans, 1984; Ono *et al.*, 1990). These two residues

Table I. Distribution of Exons Among Cytokines and Other Growth Factors[a]

Protein	Exon I	Exon II	Exon III	Exon IV	Exon V
hIL-2 (133)		−20 to 29	30 to 49	50 to 97	98 to 133
mIL-2 (133)		−20 to 43	44 to 63	64 to 112	113 to 149
mIL-3 (139)		−27 to 28	29 to 42	43 to 74	89 to 139
				75 to 88[b]	
hIL-4 (129)		−24 to 21	22 to 37	38 to 96	97 to 129
mIL-4 (120)		−20 to 24	25 to 40	41 to 91	92 to 120
hIL-5 (112)		−22 to 26	27 to 37	38 to 80	81 to 112
mIL-5 (112)		−21 to 26	27 to 37	38 to 80	81 to 112
hIL-6 (184)	−28 to −22	−22 to 42	43 to 80	81 to 129	130 to 184
mIL-6 (187)	−24 to −18	−18 to 44	45 to 82	83 to 132	133 to 187
hIL-7 (152)	−25 to −22	−21 to 23	24 to 50	51 to 94	114 to 152
				95 to 113[c]	
mIL-7 (134)	−25 to −22	−22 to 23	24 to 50	51 to 94	95 to 134
hEPO (166)	−27 to −23	−23 to 26	27 to 55	56 to 115	116 to 166
mEPO (166)	−26 to −22	−22 to 26	27 to 55	56 to 115	116 to 166
hG-CSF (174)	−30 to −17	−17 to 38	39 to 74	75 to 123	124 to 177
hGM-CSF (118)		−26 to 27	28 to 41	42 to 83	84 to 118[b]
mGM-CSF (118)		−26 to 27	28 to 41	42 to 83	84 to 118
hGH (191)	−26 to −23	−23 to 32	33 to 71	72 to 125	126 to 191
mGH (190)	−26 to −23	−23 to 30	31 to 70	71 to 124	125 to 190
hPRL (199)	−29 to −20	−19 to 39	40 to 76	77 to 136	137 to 199
mPRL (197)	−29 to −20	−19 to 37	38 to 74	75 to 134	135 to 197

[a] Exons II, III, IV, and V correspond to exons 1, 2, 3, and 4, respectively, in the text.
[b] mIL-3 and GM-CSF have one extra exon.
[c] hIL-7 has one extra exon which is missing in mIL-7.

are flanked by two polar residues and, in most cases, they are either basic (Arg and Lys) or acidic (Asp and Glu) residues (Fig. 1). The consensus sequence pattern around this Phe-Leu is $\phi Xp_1F(Y)Lp_2XX\phi p_3$, where ϕ is a hydrophobic residue, X is any amino acid, and p is a polar residue. The two nonpolar positions (ϕ) are frequently occupied by Leu and, in most cases, the last polar type (p) position is either Arg or Lys.

The second most obvious conservation occurs at the beginning of exon 3 (region 3A in Fig. 1) and again an aromatic residue (Phe, Tyr, and Trp) is present in this motif ($\phi XXF, Y/W$). The first position ϕ is occupied by a nonpolar residue, preferably a leucine. This sequence pattern is observed for 18 out of 25 cytokines. Interestingly, the aromatic residue at the fourth position is followed by another sequence motif QGL (Gln-Gly-Leu) for G-CSF, EPO, GM-CSF, and Il-5. The mouse sequence of IL-9 has the last two residues GL at the corresponding position. Another characteristic sequence pattern shared by all growth factors in the later part of exon 3 is the occurrence of hydrophobic residues in every third or fourth position for a 12-residue stretch (region 3B in Fig. 1). Though this is a common pattern for an α-helical repeat, most of the hydrophobic residues observed in this region for growth factors are Leu and Ile. Regarding the conserved region in exon 2, there is a well-conserved proline preceded by a nonpolar residue (region 2 in Fig. 1). This nonpolar residue is Leu, Ile, or Val. In many cases, nonpolar residues are also conserved at three residues downstream from proline (Fig. 1).

In addition to these consensus patterns, there are also other discrete conserved regions in each exon that are unique among growth factors. For instance, most of the growth factors start with a stretch of Pro, Ser, and Thr. In exon 1, many of the nonpolar and polar residues occur at similar positions (region 1 in Fig. 1). The polar residue are also similar types in many of these locations.

3.3. Homologous Regions Between Specific Growth Factors

Apart from these global conserved regions, we also noticed certain regions that are homologous between two specific growth factors. To identify these local homologies, we considered both identical as well as highly conservative substitutions between two sequences. Figure 2 shows these homologous blocks for different growth factors.

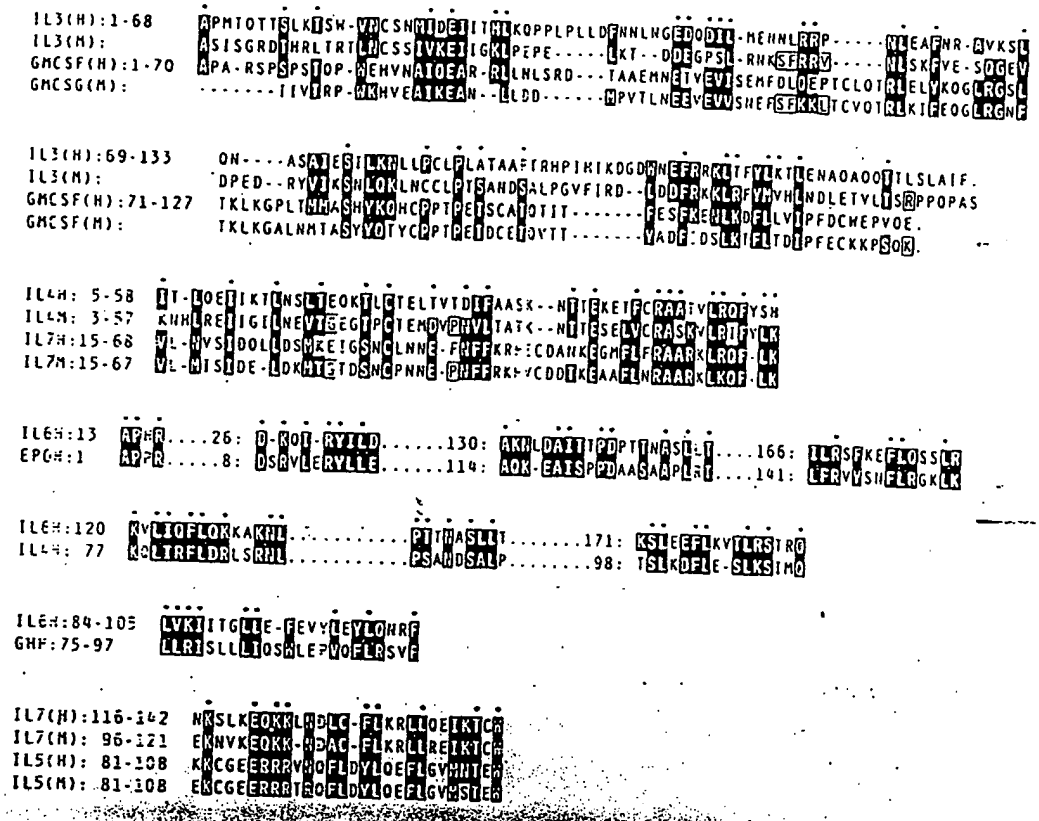

Fig. 2. Local homologies observed between various growth factors. Identical, highly conserved substitutions and residues with similar properties: in at least three out of four are shown in dark background (reverse video); between-human species are shaded; between-murine species are boxed. Asterisks denote identical residues or highly conserved substitutions.

The homology found between IL-3 and GM-CSF for the entire sequence only with five gaps is very striking. If we consider both identical and similar type residues, about 50% of human IL-3 homologous to human GM-CSF.

IL-6 shares sequence homology with a number of proteins. There are two discrete homologous regions between IL-6 and EPO in exon 4. The homology between these two growth factors is also observed near the N-terminal end of exon 1. Our analysis also indicates that the disulfide loop (Cys separated by 5 residues) observed for IL-6 at the beginning of exon 2 is probably replaced by a shorter loop (Cys separated by three residues) in the case of EPO. Growth hormone also shows some specific homology to IL-6, in particular, the conservation of aromatic residues in exon 3. These two proteins are not only similar in size but their intron/exon positions are also in close proximity. Another protein that shows specific homology to IL-6 is IL-4. The C-terminal end of exon 3 and part of exon 4 are homologous between IL-6 and IL-4.

Other cytokines also have specific homologies. For example, part of exon 3 and part of exon 4 of IL-7 are homologous to the corresponding regions in IL-4 and IL-5, respectively (Fig. 2).

3.4. Secondary Structure Analysis

The existence of discrete homologous regions among growth factors would also imply a similar secondary structural pattern for these cytokines. The secondary structural alignment confirms our prediction and reveals structural features that are common among these growth factors.

The secondary structures for the growth factors have been predicted using four different algorithms as described in the Methods section. We included IL-2 (Brandhuber *et al.*, 1987) and GH (Abdel-Meguid *et al.*, 1987) for which X-ray data are available, in our prediction analysis to check the reliability of these prediction methods. The secondary structure predictions not only indicated a significant amount of α-helical structures, but it also showed some definite β-sheet or ambiguous (either α-helical or β-sheet) structures for each cytokine, including IL-2 and GH. However, the X-ray data for IL-2 and GH indicates that the regions predicted as possible β-sheets for these two proteins are found to be either part of α-helical structures or irregular regions. A high sequence homology around these regions strongly suggests that other proteins may also have similar types of structures for these segments.

Further experimental evidence that other proteins also belong to α-helical structure family comes from the CD measurements of IL-4 (Windsor *et al.*, 1991), IL-6 (Kruttgen *et al.*, 1990), EPO (Lai *et al.*, 1986), GM-CSF (Wingfield *et al.*, 1988), and G-CSF (Lu *et al.*, 1989). Based on these experimental data, we expect that all these proteins are completely devoid of β-sheet structures and further assume that all regions predicted as definite β-sheets share α-helical structures. The location of α-helices predicted for each cytokine is given in Table II.

According to our predictions, exon 4 of all cytokines (exon 5 for IL-3) starts with an irregular structure followed by a definitive α-helical segment (Fig. 3). The highly conserved residue pair Phe-Leu forms part of the α-helical segment. This helix which is near the C-terminal end of the amino acid sequence, is the last helical segment in all cytokines included in our analysis. Exon 3 has two major α-helices separated by a small loop for all growth factors (Fig. 3). Exon 1 also has one major helix for all cytokines. Surprisingly, the major portion of exon 2 is a random structure for IL3, IL-4, IL-5, IL-7, and IL-9. Though IL-6 and G-CSF show some helical conformation in exon 2, they are also likely to have irregular structure for this region or it may not be part of the core structure, because of the occurrence of two disulfide bonds and prolines in this segment. Also, a consensus secondary structure prediction from the three homologous proteins—IL-6, G-CSF, and cMGF—precludes the presence of α-helix in this region. GM-CSF and EPO are the only proteins that seem to have a definite α-helical structure in exon 2. It is known from X-ray analysis that IL-2 also has α-helical structure in exon 2. Introns in the peptides are observed to occur either between helices or very near the periphery of α-helical regions (Fig. 3).

Table II. Prediction of α-Helices in Cytokines

Protein	α-Helical segments*
IL-6	26–45, 58–72, 87–108, 115–140, 144–170
GCSF	13–27, 45–54, 78–95, 103–126, 147–172
EPO	5–24, 46–55, 60–80, 91–114, 136–160
IL-4	5–21, 48–60, 72–88, 108–124
IL-7	10–24, 50–70, 75–91, 121–147
IL-5	8–26, 41–54, 64–80, 90–114
IL-9	4–19, 45–62, 67–82, 97–118
IL-3	9–29, 48–65, 68–82, 104–123
GM-CSF	13–26, 32–43, 53–67, 73–85, 101–119

* The sequence numbers correspond to the human version.

It could be argued that the sequence pattern observed for growth factors are characteristic of α-helical-type proteins. To validate the significance of these sequence homologies found for cytokines, we also analyzed the sequences of other α-helical-type proteins, as a negative control. The proteins we studied include hemerythrin (Ward *et al.*, 1975), cytochrome *b5* (Mathews *et al.*, 1972), cytochrome *c'* (Finzel *et al.*, 1985), and cytochrome *b562* (Lederer *et al.*, 1981). None of these proteins showed the characteristic sequence patterns observed for the growth factors.

3.5. Folding Topologies

The three-dimensional structures for IL-2 (Brandhuber *et al.*, 1987) and GH (Abdel-Meguid *et al.*, 1987) determined by X-ray diffraction show that both have left-handed antiparallel four α-helical bundle-type structures. However, the connectivity of α-helices for these two proteins is different. IL-2 has the typical up-down-up-down connectivity with helix A connected to B, B to C, and C to D (Fig. 4a). The α-helices of GH are joined by up-up-down-down connectivity (one short distance connection and two long distance or overhand connections). The topological associations of the four α-helices are A to C, C to B, and B to D (Fig. 4b).

Crystal structure data analysis of four-α-helical type proteins indicates that 12 out of 13 α-helical proteins, including IL-2 and GH, adopt all-antiparallel α-helical bundles (Presnell and Cohen, 1989).

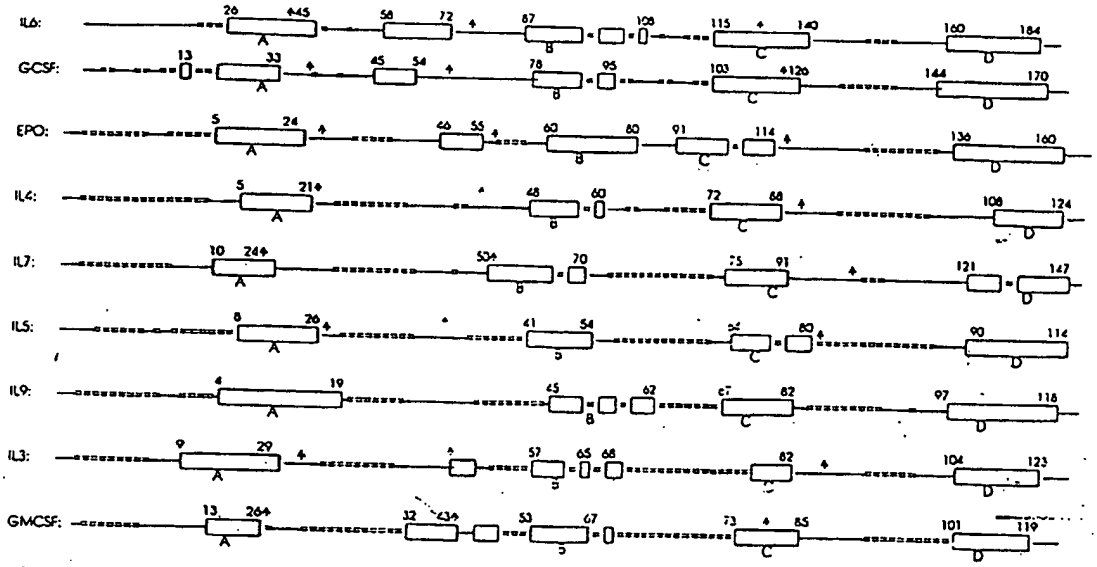

Fig. 3. Comparison of secondary structures for growth factors. The initial alignment is taken from Fig. 1. Cylinders are α-helices and solid lines are other structures. Broken lines represent gaps. Arrows indicate intron/exon boundaries.

Theoretical energy calculations also suggest that all-antiparallel-type is more stable than the parallel type (Chou *et al.*, 1988). Hence, it is a reasonable assumption that all cytokines will adopt the all-antiparallel-type structures. There are three possible topological connectivities for left-handed all-antiparallel bundles. These connectivities are up-down-up-down, up-up-down-down, and up-down-down-up (Fig. 4a–c). GH and IL-2 represent the first two topologies, respectively.

One common feature that is shared by GH and IL-2 is that the middle two α-helices of the four α-helical bundles are connected by a short loop and are antiparallel to each other. Since our prediction analysis also indicates that exon 3 of many cytokines has two α-helices separated by a short distance, they would invariably be antiparallel to each other and form the middle two α-helical segments. This would eliminate one of the three possibilities (up-down-down-up), leaving the remaining two topological connections as the possible candidates (Fig. 4). Using the disulfide bond assignments as a further guidance, it should be possible to predict the folding topologies of certain proteins.

3.5.1. IL-4

Human and murine IL-4 have three disulfide bonds (Carr *et al.*, 1991). While the second S-S bond is identical for both species, the locations of the other two S-S bonds are different. In human IL-4, the first Cys is bonded to the sixth Cys, which is located near the C-terminal end. For murine IL-4, the same first Cys is bonded to the sixth Cys located in a loop between helices C and D (Fig. 4d and e). If IL-4s of both the species adopt the same topology, then the location of these cysteines for both murine and human IL-4 should be on the same side. This condition will be satisfied only when helices C and D are parallel to each other and helices A and D are antiparallel (i.e., the topology of IL-4 will have the up-up-down-down connectivity similar to GH folding).

3.5.2. GM-CSF and IL-3

GM-CSF is another protein for which disulfide arrangements are known (Lu *et al.*, 1989; Shanafelt and Kastelein, 1989). It has two disulfide bonds and both are located in the second part of the molecule. The first S-S bond (S-S bond between first and third cysteine) connects the helical segment B and the loop between helices C and D, while the second S-S bond (S-S bond between second and fourth cysteine) connects the loop between C and D and the C terminal end of the D-helix (Fig. 4c). This arangement is possible only if helices C and D are parallel with one overhand connection. The orientation of helix A relative to helix B is not clear. There are two potential

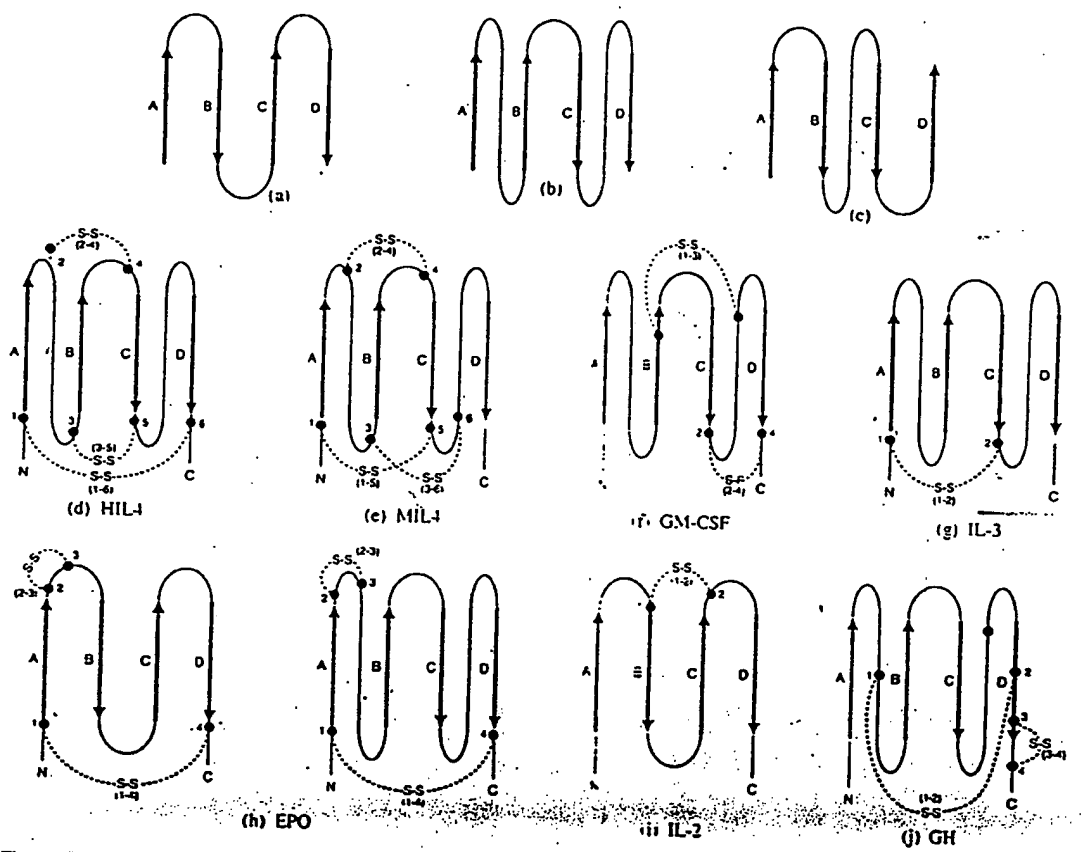

Fig. 4. Schematic representation of three possible topological connectivities for an all-antiparallel four α-helical bundles. Each connectivity can be either left-handed or right-handed. (a) up-down-up-down. IL-2 belongs to this category; (b) up-up-down-down with two hand-over connections, GH belongs to this type; (c) up-down-down-up with one hand-over connection in the middle; disulfide bond arrangements and the topological connections for human (d) IL-4 and murine (e) IL-4; (f) GM-CSF; (g) IL-3; (h) two possible connections for EPO, (i) IL-2, and (j) GH. Thick lines (A, B, C and D) and thin lines represent α-helices and other structures, respectively. See Fig. 3 for the locations of α-helical segments. The location of Cys along the sequence is represented by the sequential numbers 1, 2, 3, etc. The solid circles denote cysteines.

helical segments preceding helix B of GM-CSF (Fig. 3). If this protein is also all-antiparallel-type, then one of these two helices should be parallel to helix B. The topological prediction for the latter part of the molecule agrees with the models proposed by Kaushansky *et al.* (1990) and Parry *et al.* (1988). However, both models do not represent all-antiparallel-type structures. IL-3 has one disulfide bond (Clark-Lewis *et al.*, 1987) that connects the putative N-terminus of helix A and the putative C-terminus of helix C, which would argue that these two helices are antiparallel to each other (Fig. 4g). If GM-CSF is also structurally homologous to IL-3, this would further support that helix A of GM-CSF is parallel to helix B of the four α-helical bundle. A similar model has been proposed by Lokker *et al.* (1991) for IL-3 which has the up-up-down-down connectivity.

3.5.3. *EPO*

For EPO, there is a disulfide bond between the N-terminal end and the C-terminal end (McDonald *et al.*, 1986). This would suggest that these two α-helical segments are oriented in opposite directions (Fig. 4h). Since the middle two α-helices are also predicted to be antiparallel to each other, this protein can adopt any one of the two possibilities (either up-down-up-down or up-up-down-down folding topology) to satisfy the disulfide bonding connection (Fig.

4h). In this regard, Bazan (1990) has proposed the latter topology for EPO.

3.5.4. Other Proteins

The disulfide assignments for IL-5, IL-7, and IL-9 are not known. IL-5 has only two cysteines which are located before and after helix B and helix C, respectively. If we assume a disulfide bond between these two Cys, it would orient helices B and C in opposite directions, again supporting our prediction that the central two α-helices of a four α-helical bundle are antiparallel. Finally, the loop sizes and the α-helices of IL-6, G-CSF, and c-MGF are comparable to GH, which tempt us to speculate that their structures are also similar to GH in that they all have up-up-down-down connectivity. Though this is consistent with the model proposed by Bazan (1990), more data is necessary to confirm our prediction. However, the observation of discrete blocks of homologous amino acid sequences shared by many cytokines (Table II) would strongly suggest that IL-4, IL-5, IL-6, IL-7, EPO, G-CSF, and c-MGF all have similar folding topologies.

4. DISCUSSION

In this study we have shown that the sequences of nine different cytokines and growth factors can be aligned and their secondary structure predicted. We also examined several other cytokines and growth factors whose receptors do not belong to the cytokine receptor superfamily (Bazan, 1990; Goodwin *et al.*, 1990). For example, we attempted to match the sequence of IL-1 (Eisenberg *et al.*, 1991) with the alignment reported here without success, indicating that IL-1 does not belong to this ligand superfamily. Interestingly, IL-1 has been shown to be primarily β-sheet in structure (Finzel *et al.*, 1989; Graves *et al.*, 1990). We also examined the sequence of IL-11 (Paul *et al.*, 1990). Its sequence does not match our alignment either and we would conclude that this cytokine does not belong to the ligand superfamily identified in this study. As a corollary to this conclusion, our work would predict that the receptor for IL-11 would not belong to the cytokine receptor superfamily. The identification and sequencing of the IL-11 receptor will be required to test our prediction.

The primary sequence and secondary structure analysis of cytokines and other growth factors examined in our study suggest that these proteins might have evolved from a single primordial gene. The features that are found common at genomic, primary, and secondary structural levels further support this hypothesis. However, the possibility that certain proteins within this family are likely to adopt different folding topologies would also imply at least two subclasses within this large superfamily.

The receptors for the cytokines studied here belong to a superfamily that also may have evolved from a single primordial gene. It is tempting to speculate upon the hypothesis of a primordial ligand and a primordial receptor divergently coevolving into the family of ligands and receptors discussed here. A detailed comparison of the subfamily pattern within both ligand and receptor superfamilies might allow us to test the validity of this hypothesis.

Our analysis also suggests that overlapping functions of certain proteins can be attributed to functional residues found at similar locations due to similar folding of the entire molecule or a major portion of it. In this context, IL-3 and GM-CSF share certain biological activities and they can cross-compete for binding on some cell lines. For example, the regions (18–22, 34–41, 52–61, and 94–115) of GM-CSF have recently been shown to be responsible for its biological activity (Shanafelt *et al.*, 1991). In our study (Fig. 2), we show that it is these regions of GM-CSF that are also homologous to IL-3 (Fig. 2). In addition, the residues Lys 110 and Leu 111 of human IL3, which have been demonstrated to be critical for its biological activity (Lokker *et al*, 1991), overlap in our alignment with the functional region 94–115 of GM-CSF (Figs. 2 and 3). Subtle differences within these functional regions might also dictate the differences in biological activity between two growth factors. This could obviously be tested by site-specific mutagenesis followed by a ligand-binding assay. Recently, this kind of approach has been attempted for GH to induce prolactin activity, and the results have been very successful (Cunningham *et al.*, 1990: Cunningham and Wells, 1991). Our sequence alignment and similar folding patterns inferred from these results should be extremely valuable in identifying critical regions within other members of this ligand family and in modifying a growth factor for making it highly selective for a particular function.

ACKNOWLEDGMENTS

We thank Dr. John McPherson for helpful discussions and encouragement.

NOTE ADDED IN PROOF

The crystal structure of GM-CSF has been published (Diederichs, K., Boone, T. and Karplus, P.A. (1991) *Science*, vol. 254, 1779–1782) after the submission of this manuscript. The location of α-helices (A: 13–28, B: 55–64, C: 74–87 and D: 103–116) and their up-up-down-down connectivity agree very well with our prediction.

REFERENCES

Abdel-Meguid, S. S., Shieh, H.-S., Smith, W. W., Dayringer, H. E., Violand, B. N., and Bentle, L. A. (1987). *Proc. Natl. Acad. Sci. USA* 84, 6434–6437.

Arai, K.-I., Lee, F., Miyajima, A., Miyatake, S., Arai, N., and Yokota, T. (1990). *Annu. Rev. Biochem.* 59, 783–836.

Azuma, C., Tanabe, T., Konishi, M., Kinashi, T., Noma, T., Matsuda, F., Yaoita, Y., Takatsu, K., Hammarstrom, L., Smith, C. I. E., Severinson, E., and Honjo, T. (1986). *Nucl. Acids Res.* 14, 9149–9158.

Barta, A., Richards, R., Baxter, J. D., and Shine, J. (1981). *Proc. Natl. Acad. Sci. USA* 78, 4853–4871.

Bazan, J. F. (1990). *Immunol. Today* 11, 350–354.

Bazan, J. F. (1989). *Biochem. Biophys. Res. Comm.* 164, 788.

Brandhuber, B. J., Boone, T., Kenney, W. C., and McKay, D. B. (1987). *Science* 238, 1707–1709.

Carr, C., Aykent, S., Kimack, N. M., and Levine, A. D. (1991). *Biochemistry* 30, 1515–1523.

Chou, P.-Y., and Fasman, G. D. (1978). *Annu. Rev. Biochem.* 47, 251–276.

Chou, K.-C., Maggiora, G. M., Nemethy, G., and Scheraga, H. A. (1988). *Proc. Natl. Acad. Sci. USA* 85, 4295–4299.

Clark-Lewis, I., Lopez, A. F., Vadas, M., Schrader, J. W., Hood, L. E., and Kent, S. B. H. (1987). In *Molecular Basis for Lymphokine Action* (Webb, D., Pierce, C., and Cohen, S.), Humana, Clifton, New Jersey, pp. 339–351.

Cohen, F. E., Abarbanel, R. M., Kuntz, I. D., and Fletterick, R. J. (1986). *Biochemistry* 25, 266–275.

Cooke, N. E., Coit, D., Shine, J., Baxter, J. D., and Martial, J. A. (1981). *J. Biol. Chem.* 256, 4007–4016.

Cunningham, B. C., Henner, D. J., and Wells, J. A. (1990). *Science* 247, 1461–1465.

Cunningham, B. C., and Wells, J. A. (1991). *Proc. Natl. Acad. Sci. USA* 88, 3407–3411.

D'Andrea, A. D., Fasman, G. D., and Lodish, H. F. (1989). *Cell* 58, 1023–1024.

DeNoto, F. M., Moore, D. D., and Goodman, H. M. (1981). *Nuc. Acids. Res.* 9, 3719–3730.

Devereux, J. R., Haeberli, P., and Smithies, O. (1984). *Nucl. Acids Res.* 12, 387–395.

Eisenberg, S. P., Brewer, M. T., Verderber, E., Heimdal, P., Brandhuber, B. J., and Thompson, R. C. (1991). *Proc. Natl. Acad. Sci. USA* 88, 5232–5236.

Finzel, B. C., Clancy, L. L., Holland, D. R., Muchmore, S. W., Watenpaugh, K. D., and Einspahr, H. M. (1989). *J. Mol. Biol.* 209, 779–791.

Finzel, B. C., Weber, P. C., Hardman, K. D., and Salemme, F. R. (1985). *J. Mol. Biol.* 186, 627–643.

Fung, M. C., Hapel, A. J., Ymer, S., Cohen, D. R., Johnson, R. M., Campbell, H. D., and Young, I. G. (1984). *Nature* 307, 233–237.

Garnier, J., Osguthorpe, D. J., and Robson, B. (1978). *J. Mol. Biol.* 120, 97–120.

Goodwin, R. G., Friend, D., Ziegler, S. F., Jerzy, R., Falk, B. A., Gimpel, S., Cosman, D., Dower, S. K., March, C. J., Namen, A. E., and Park, L. S. (1990). *Cell* 60, 941–951.

Goodwin, R. G., Lupton, S., Schmierer, A., Hjerrild, K. J., Jerzy, R., Clevenger, W., Gillis, S., Cosman, D., and Namen, A. E. (1989). *Proc. Natl. Acad. Sci. USA* 86, 302–306.

Gough, N. M., Gough, J., Metcalf, D., Kelso, A., Grail, D., Nicola, N. A., Burgess, A. W., and Dunn, A. R. (1984). *Nature* 309, 763–767.

Graves, B. J., Hatada, M. H., Hendrickson, W. A., Miller, J. K., Madison, V. S., and Satow, Y. (1990). *Biochemistry* 29, 2679–2684.

Hirano, T., Yasukawa, K., Harada, H., Taga, T., Watanabe, Y., Matsuda, T., Kashiwamura, S.-I., Nakajima, K., Koyama, K., Iwamatsu, A., Tsunasawa, S., Sakiyama, F., Matsui, H., Takahara, Y., Tanaguchi, T., and Kishimoto, T. (1986). *Nature* 324, 73–76.

Jacobs, K., Shoemaker, C., Rudersdorf, R., Neill, S. D., Kaufmann, R. J., Mufson, A., Seehra, J., Jones, S. S., Hewick, R., Fritsch, E. F., Kawakita, M., Shimizu, T., and Miyake, T. (1985). *Nature* 313, 806–810.

Kashima, N., Nishi-Takaoka, C., Fujita, T., Taki, S., Yamada, G., Hamuro, J., and Taniguchi, T. (1985). *Nature* 313, 402–404.

Kaushansky, K., Brown, C. B., and O'Hara, P. J. (1990). *Int. J. Cell Cloning* 8 (Suppl. 1), 26–34.

Kinashi, T., Harada, N., Severinson, E., Tanabe, T., Sideras, P., Konishi, M., Azuma, C., Tominaga, A., Bergstedt-Lindqvist, S., Takahashi, M., Matsuda, F., Yaoita, Y., Takatsu, K., and Honjo, T. (1986). *Nature* 324, 70–73.

Krutgen, A., Rose-John, S., Moller, C., Wroblowski, B., Wollmer, A., Mullberg, J., Hirano, T., Kishimoto, T., and Heinrich, P. (1990). *FEBS Lett.* 262, 323–326.

Lai, P.-H., Everett, R., Wang, F.-F., Arakawa, T., and Goldwasser, E. (1986). *J. Biol. Chem.* 261, 3116–3121.

Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D., and Mathews, F. S. (1981). *J. Mol. Biol.* 148, 427–448.

Lee, F., Yokota, T., Otsuka, T., Meyerson, P., Villaret, D., Coffman, R., Mosmann, T., Rennick, D., Rochm, N., Smith, C., Zlotnik, A., and Arai, K.-I. (1986). *Proc. Natl. Acad. Sci. USA* 83, 2061–2065.

Leutz, A., Damm, K., Sterneck, E., Kowenz, E., Ness, S., Frank, R., Gausepohl, H., Pan, Y.-C. E., Smart, J., Hayman, M., and Graf, T. (1989). *EMBO J.* 8, 175–181.

Lim, V. I. J. (1974). *J. Mol. Biol.* 88, 857–894.

Lin, F.-K., Suggs, S., Lin, C.-H., Browne, J. K., Smalling, R., Egrie, J. C., Chen, K. K., Fox, G. M., Martin, F., Stabinsky, Z., Badraw, S. M., Lai, P.-H., and Goldwasser, E. (1985). *Proc. Natl. Acad. Sci. USA* 82, 7580–7584.

Linzer, D. I. H., and Nathans, D. (1984). *Proc. Natl. Acad. Sci. USA* 81, 4255–4259.

Lokker, N. A., Zenke, G., Strittmatter, U., Fagg, B., and Movva, N. R. (1991). *EMBO J.* 10, 2125–2131.

Lu, H. S., Boone, T. C., Souza, L. M., and Lai, P.-H. (1989). *Arch. Biochem. Biophys.* 268, 81–92.

Lupton, S. D., Gimpel, S., Jerzy, R., Brunton, L. L., Hjerrild, K. A., Cosman, D., and Goodwin, R. G. (1990). *J. Immunol.* 144, 3592–3601.

Mathews, F. S., Argos, P., and Levine, M. (1972). *Cold Spring Harbor Symp. Quant. Biol.* 36, 387–395.

McDonald, J. D., Lin, F.-K., and Goldwasser, E. (1986). *Mol. Cell. Biol.* 6, 842–848.

Metcalf, D. (1989). *Nature* 339, 27–30.

Mizel, S. B. (1989). *FASEB J.* 3, 2379–2388.

Nagata, S., Tsuchiya, M., Asano, S., Kaziro, Y., Yamazaki, T., Yamamoto, O., Hirata, Y., Kubota, N., Oheda, M., Nomura, H., and Ono, M. (1986). *Nature* 319, 415–418.

Nicola, N. A. (1989). *Ann. Rev. Biochem.* 58, 45–77.

Nicola, N. A., and Metcalf, D. (1991). *Cell* 67, 1–4.

Ono, M., Takayama, Y., Rand-Weaver, M., Sakata, S., Yasunaga, T., Noso, T., and Kawauchi, H. (1990). *Proc. Natl. Acad. Sci. USA* 87, 4330–4334.

Page, G. S., Smith, S., and Goodman, H. M. (1981). *Nuc. Acid. Res.* 9, 2087–2104.

Parry, D. A. D., Minasian, E., and Leach, S. J. (1988). *J. Mol. Recogn.* 1, 107–110.

Paul, S. R., Bennett, F., Calvetti, J. A., Kelleher, K., Wood, C. R., O'Hara, Jr., R. M., Leary, A. C., Sibley, B., Clark, S. C., Williams, D. A., and Yang, Y.-C. (1990). *Proc. Natl. Acad. Sci. USA* 87, 7512–7516.

Paul, W. E. (1989). *Cell* 57, 521–524.

Presnell, S. R., and Cohen, F. E. (1989). *Proc. Natl. Acad. Sci. USA* 86, 6592–6596.

Shanafelt, A. B., Johnson, K. E., and Kastelein, R. A. (1991). *J. Biol. Chem.* 266, 13,804–13,810.

Shanafelt, A. B., and Kastelein, R. A. (1989). *Proc. Natl. Acad. Sci. USA* 86, 4872–4876.

Taniguchi, T., Matsui, H., Fujita, T., Takaoka, C., Kashima, N., Yoshimoto, R., and Hamuro, J. (1983). *Nature* 302, 305–310.

Tsuchiya, M., Asano, S., Kaziro, Y., and Nagata, S. (1986). *Proc. Natl. Acad. Sci. USA* 83, 7633–7637.

Van Snick, J., Cayphas, S., Szikora, J.-P., Renauld, J.-C., Van Roost, E., Boon, T., and Simpson, R. J. (1988). *Eur. J. Immunol.* 18, 193–197.

Ward, K. B., Hendrickson, W. A., and Klippenstein, G. L. (1975). *Nature* 257, 818–821.

Windsor, W. T., Syto, R., Le, H. V., and Trotta, P. P. (1991). *Biochemistry* 30, 1259–1264.

Wingfield, P., Graber, P., Moonen, P., Craig, S., and Pain, R. H. (1988). *Eur. J. Biochem.* 173, 65–72.

Wong, G. G., Witek, J. S., Temple, P. A., Wilkens, K. M., Leary, A. C., Luxenberg, D. P., Jones, S. S., Brown, E. L., Kay, R. M., Orr, E. C., Shoemaker, C., Golde, D. W., Kaufman, R. J., Hewick, R. M., Wang, E. A., and Clark, S. C. (1985). *Science* 228, 810–815.

Yang, Y.-C., Ciarletta, A. B., Temple, P. A., Chung, M. P., Kovacic, S., Witek-Giannotti, J. S., Leary, A. C., Kriz, R., Donahue, R. E., Wong, G. G., and Clark, S. C. (1986). *Cell* 47, 3–10.

Yang, Y.-C., Ricciardi, S., Ciarletta, A., Calvetti, J., Kelleher, K., and Clark, S. C. (1989). *Blood* 74, 1880–1884.

Yokota, T., Otsuka, T., Mosmann, T., Banchereau, J., DeFrance, T., Blanchard, D., DeVries, J. E., Lee, F., and Arai, K.-I. (1986). *Proc. Natl. Acad. Sci. USA* 83, 5894–5898.